



Une méthode de binarisation hiérarchique floue

Salvatore Tabbone, Thi Oanh Nguyen, Gérald Masini

► To cite this version:

Salvatore Tabbone, Thi Oanh Nguyen, Gérald Masini. Une méthode de binarisation hiérarchique floue. Colloque International Francophone sur l'Écrit et le Document - CIFED 2006, Sep 2006, Fribourg, Suisse. pp.43-48. hal-00118323

HAL Id: hal-00118323

<https://hal.science/hal-00118323>

Submitted on 11 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une méthode de binarisation hiérarchique floue

Salvatore Tabbone¹ – Thi-Oanh Nguyen^{1,2} – Gérald Masini¹

¹ LORIA - CNRS - Université de Nancy 2

Campus Scientifique, B. P. 239, 54506 Vandœuvre-lès-Nancy Cedex

² Institut de la Francophonie pour l'Informatique (IFI)

42 Ta Quang Buu, Hai Ba Trung, Hanoï, Viêt-Nam

Salvatore.Tabbone@loria.fr, Thi-Oanh.Nguyen@loria.fr, Gerald.Masini@loria.fr

Résumé : *Cet article propose une approche hiérarchique pour la binarisation d'images de documents graphiques, basée sur un découpage en arbre quaternaire, dont l'intérêt est de traiter l'image à différents niveaux. Une image est d'abord segmentée à bas niveau grâce à une approche globale de binarisation, puis le résultat de cette première segmentation est localement affiné après un découpage récursif de l'image initiale. Ce découpage hiérarchique permet de combiner les avantages de deux types d'approches duales, globale et locale. Les résultats sont prometteurs : selon des critères d'homogénéité et de contraste fixés, la séparation entre le fond et le contenu du document est meilleure que celle obtenue avec certaines méthodes classiques.*

Mots-clés : binarisation hiérarchique, arbre quaternaire, sous-ensembles flous, fonction d'appartenance

1 Introduction

L'une des premières étapes d'un système traitant un document graphique, voire la première, est la binarisation, qui consiste à segmenter le document en deux classes, le fond, d'une part, et l'objet, c'est-à-dire les autres composantes (texte, graphique...), d'autre part. Elle permet de réduire de manière appréciable la quantité d'informations à manipuler. La qualité du résultat obtenu est cruciale puisqu'elle conditionne fortement celle des étapes postérieures de la chaîne de traitements appliqués au document.

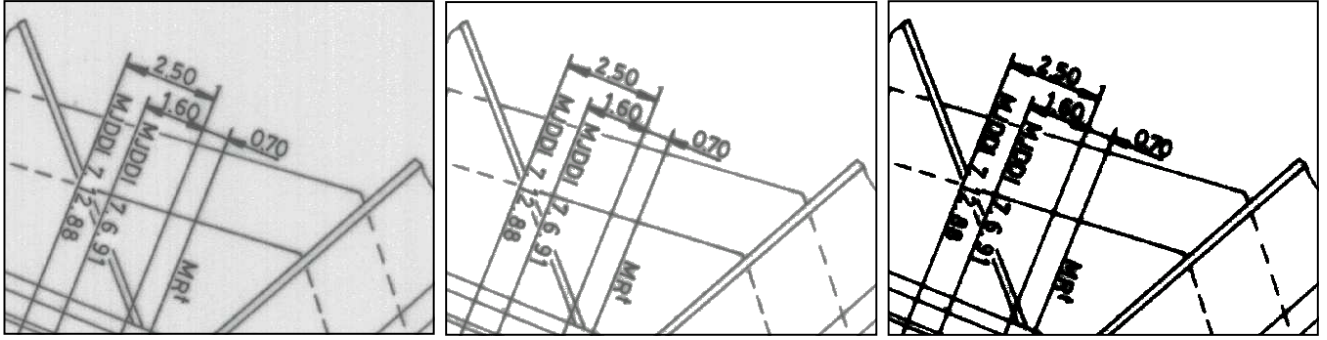
Les méthodes de binarisation se répartissent en trois grandes catégories. Les approches globales utilisent un seuil de coupure déterminé globalement à partir de la distribution photométrique des pixels, par optimisation d'un critère comme la variance ou l'entropie [OTS 79, KUR 92, CHE 98, TAO 02, TAB 03]. Ce seuil est ensuite uniformément appliqué à tous les pixels de l'image, pour diviser l'histogramme des niveaux de gris en deux classes correspondant au fond et à l'objet, sous l'hypothèse de bimodalité de l'histogramme et/ou de normalité des distributions. Ces méthodes donnent donc de bons résultats lorsque l'image contient des classes qui vérifient une certaine homogénéité, traduite par un histogramme multimodal. Elles sont généralement conçues pour les images à niveaux de gris.

Si l'image est relativement dégradée et la répartition des niveaux de gris non uniforme, un seuil global unique ne permet pas de séparer efficacement le fond de l'objet. Il faut alors tenir compte localement de la répartition spa-

tiale et photométrique des pixels. L'image est donc découpée en régions, soit de même taille (dans le cadre d'un découpage systématique), soit de tailles différentes (en s'appuyant par exemple sur une segmentation en contours) [PAR 91, JAI 92], puis une répartition statistique des pixels [EIK 91] ou du contraste [KAM 93, YAN 00] est calculée localement pour chacune des régions. De nombreuses approches de cette catégorie s'appuient sur les travaux de O'Gorman [O'G 94] qui a proposé en 1994 l'histogramme de répartition de connectivité des pixels (*run length histogram*), dont les modes donnent des informations sur les épaisseurs des traits graphiques contenus dans le document.

Enfin, les approches hybrides font coopérer des méthodes locales et globales, par exemple en fonction du type du contenu du document, texte ou graphique [SAU 00]. Une technique intéressante consiste à déterminer localement une région modèle dont les caractéristiques sont ensuite utilisées pour traiter l'image entière [DAW 02]. Dans le même ordre d'idée, une décomposition en ondelettes de Hadamard permet d'estimer automatiquement l'épaisseur du trait des caractères d'un document écrit en chinois avant une segmentation globale [CHA 99].

Malgré le grand nombre de méthodes existantes, très peu sont adaptées à la problématique du document graphique. Hormis la méthode de Otsu [OTS 79] qui sert souvent de prétraitement pour la binarisation de documents [LIU 97, CHA 99], notons celle de Trier et Taxt [TRI 95], qui s'appuie sur la polarité du Laplacien pour segmenter l'image en régions, et celle de Kim [KIM 02], qui utilise une technique de ligne de partage des eaux pour segmenter des documents de type texte. En outre, les techniques reposant sur la détermination de la largeur du trait des caractères ne s'appliquent pas toujours aux documents graphiques, qui peuvent parfois ne contenir qu'une seule composante connexe. En fait, c'est la proximité entre certaines petites composantes textuelles et les (grandes) composantes graphiques qui est la source de la principale difficulté. Les méthodes dites globales sont robustes au bruit mais ne permettent pas de bien séparer les petites composantes des grandes. À l'inverse, les approches locales sont plus sensibles au bruit mais donnent de meilleurs résultats dans la séparation des composantes. Les approches hybrides, et plus particulièrement celles qui sont multirésolutions ou multiniveaux, nous semblent les mieux appropriées au traitement des documents graphiques. Nous nous sommes



(a) image originale

(b) image seuillée

(c) image binaire correspondante

FIG. 1 – Première étape : le seuillage global.

donc intéressés aux techniques à base d'arbre quaternaire (*quadtree*) qui permettent de séparer le fond de ce qui constitue le document proprement dit grâce à des tests statistiques.

Plus précisément, nous avons amélioré l'approche de Gadi et Benslimane [GAD 00], trop sensible au bruit de fond de l'image, en perfectionnant l'étape d'initialisation. Le paragraphe 2 présente le principe de la méthode et le paragraphe 3 les résultats expérimentaux, accompagnés d'une comparaison avec certaines des méthodes de segmentation les plus usuelles. Le paragraphe 4 donne brièvement les perspectives du travail décrit en guise de conclusion.

2 Principe de la binarisation

La méthode proposée, qui peut être considérée comme la combinaison d'une approche globale et d'une approche locale, comprend deux étapes. La première est un seuillage global, qui élimine la plupart du fond pour garder la partie significative contenant l'objet. La seconde étape est une amélioration de la méthode locale adaptative de Gadi et Benslimane [GAD 00], déjà citée, qui affine l'image résultant de l'étape précédente pour la rendre « plus nette ».

2.1 Le seuillage global

Le seuillage global permet de définir automatiquement un seuil pour toute l'image. En principe, n'importe quelle méthode peut convenir, mais nous avons opté pour la méthode de Otsu [OTS 79], qui s'est révélée performante sur des images bimodales [SAH 88, TRI 95] et qui ne nécessite la connaissance d'aucun seuil préalable. Elle consiste à maximiser un critère de variance entre le fond et l'objet, défini par :

$$s_{optimal} = \arg \max_{t \in [min, max]} \eta(t) = \frac{\delta_B^2(t)}{\delta_T^2(t)}$$

Comme le montre la figure 1, l'image du document (cf. Fig. 1(a)) est divisée en deux classes sur la base de ce seuil optimal. Les pixels du fond sont mis à zéro et ceux désignés comme faisant partie de l'objet conservent leur valeur initiale de niveau de gris. L'image ainsi obtenue (cf. Fig. 1(b)) contient bien la partie du document qui

nous intéresse, autrement dit les parties graphiques. Certaines d'entre elles ne sont cependant pas vraiment nettes ou clairement distinguées les unes des autres car des pixels qui auraient dû être affectés au fond leur sont attribués. L'image binaire correspondante (cf. Fig. 1(c)) met bien le phénomène en évidence en le rendant plus visible. Les pixels incriminés appartiennent généralement aux frontières entre fond et objet, où l'illumination n'est pas constante.

2.2 Le seuillage adaptatif

Cette seconde étape a pour but de séparer plus finement les pixels de l'image à niveaux de gris résultante (cf. Fig. 1(b)) grâce à un seuillage adaptatif tenant compte d'informations locales au voisinage des pixels frontières entre fond et objet. La méthode retenue repose sur la théorie des sous-ensembles flous. En utilisant un critère statistique d'homogénéité, l'image est décomposée récursivement en sous-images de plus petite taille, qui sont intégrées dans une structure d'arbre quaternaire. Le processus comprend trois étapes :

1. construction de l'arbre quaternaire proprement dit,
2. calcul du degré d'appartenance au fond et à l'objet de chaque pixel, à chaque niveau de l'arbre,
3. décision finale de l'appartenance des pixels, au fond ou à l'objet.

2.2.1 Construction de l'arbre quaternaire

L'image initiale I (cf. Fig. 1(b)) est hiérarchiquement divisée en régions de tailles décroissantes formant un arbre quaternaire. La racine de l'arbre, au niveau 0, représente l'image complète. Au niveau n , un nœud représente une région d'une image de niveau $n-1$. Si cette région ne satisfait pas un critère d'homogénéité donné, elle est elle-même divisée en quatre parties carrées (ou rectangulaires) égales représentées par quatre nœuds de niveau $n+1$. Si la région satisfait le critère d'homogénéité, elle n'est plus divisée et constitue une feuille de l'arbre. Le processus de division s'arrête lorsqu'il n'y a plus de région susceptible d'être divisée.

À chaque étape, on calcule la différence entre la moyenne des pixels de la région mère et celles de ses quatre filles, ainsi que la différence entre la variance de la région mère

et celles de ses quatre filles. Classiquement, la région mère n'est pas découpée si aucune de ces différences n'est significative. Pour éviter le problème du choix du seuil permettant d'établir qu'une « différence est significative », un test statistique de Fisher est utilisé comme critère d'arrêt [GAD 00]. Il permet de comparer les moyennes et les écarts-types entre une région mère et ses quatre filles et repose sur les deux hypothèses suivantes :

$$\begin{cases} H_0 \text{ (nulle)} & \sigma = \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 \text{ et} \\ & m = m_1 = m_2 = m_3 = m_4 \\ H_1 \text{ (alternative)} & \exists j \in \{1, 2, 3, 4\} \mid \sigma_j \neq \sigma \text{ ou } m_j \neq m \end{cases}$$

où σ et σ_i sont respectivement les écarts-types calculés sur la région et ses quatre filles, et m et m_i sont respectivement les moyennes calculées sur la région et ses quatre filles. Nous supposons que les quatre régions filles sont indépendantes et présentent des distributions de niveaux de gris normales et identiques. Dans ces conditions, le test d'homogénéité f suit une distribution de Fisher $\mathcal{F}_{l;p-l-1}^\alpha$ définie par :

$$f = \frac{P \sum_{i=1}^4 \frac{(m_i - m)^2}{3}}{\sum_{i=1}^4 \sum_{j=1}^P \frac{(g_{ij} - m_i)^2}{4(P-1)}}$$

où P est le nombre de pixels de chaque région fille, g_{ij} la valeur du j^e pixel de la région fille i , l le degré de liberté (ici $l = 3$), p le nombre total de pixels de la région mère (ici $4P$) et α le niveau de confiance. Les valeurs de la distribution sont définies à partir d'une table de Fisher et la décision d'homogénéité d'une région dépend de la comparaison entre f et $\mathcal{F}_{l;p-l-1}^\alpha$. Si $f \leq \mathcal{F}_{l;p-l-1}^\alpha$, l'hypothèse H_0 dite « nulle » est vérifiée et la région est homogène. Sinon, l'hypothèse H_1 est vérifiée et la région doit encore être divisée.

Pour qu'un test statistique soit significatif, il faut évidemment que la taille de l'échantillon soit suffisamment grande. Deux conditions commandent donc en fait l'arrêt de la division d'une région : soit la taille d'une région fille est trop petite (nous l'avons fixée à 40 pixels, comme le recommandent les manuels de statistiques), soit cette taille est suffisante mais le test d'homogénéité est vérifié.

2.2.2 Appartenance au fond et à l'objet

Quand une région n'est pas homogène, ses pixels sont répartis en deux sous-ensembles flous F et O représentant respectivement le fond et l'objet, après calcul de degrés respectifs d'appartenance μ_F et μ_O à l'aide de la fonction classique de Zadeh (S_Z). Pour un pixel de niveau de gris g :

$$\begin{cases} \mu_F(x, y) = S_Z(g(x, y); a, b, c) \\ \mu_O(x, y) = Z_Z(g(x, y); a, b, c) = 1 - S_Z(g(x, y); a, b, c) \\ b = \frac{(a+c)}{2} \end{cases} = \begin{cases} 0 & g(x, y) \leq a \\ 2 \left(\frac{g(x, y) - a}{c - a} \right)^2 & a < g(x, y) \leq b \\ 1 - 2 \left(\frac{g(x, y) - c}{c - a} \right)^2 & b < g(x, y) \leq c \\ 1 & c < g(x, y) \end{cases}$$

a , b et c désignent des propriétés locales des régions. Plus précisément, l'intervalle $[m - \sigma, m + \sigma]$ est considéré comme la bande d'incertitude. Le degré d'appartenance est absolu dans toute la bande des niveaux de gris excepté l'intervalle flou. Le degré d'appartenance d'un pixel d'une région R de niveau n dans l'arbre quaternaire est donc déterminé par :

$$\begin{cases} \mu_F^n(x, y) = S_Z(g_R(x, y); m - \sigma, m, m + \sigma) \\ \mu_O^n(x, y) = 1 - \mu_F^n(x, y) \end{cases}$$

m et σ désignant respectivement la moyenne et l'écart-type de la région R .

2.2.3 Classement final des pixels

Pour décider à quel ensemble, F le fond ou O l'objet, appartient chaque pixel, un degré d'appartenance global est calculé en fonction des degrés d'appartenance obtenus aux différents niveaux de l'arbre, à l'aide d'un opérateur h . Pour l'appartenance à l'objet :

$$\mu_O^f(x, y) = h(\mu_O^1(x, y), \mu_O^2(x, y), \dots, \mu_O^N(x, y))$$

où N représente le nombre de niveaux dans lesquels apparaît le pixel concerné. L'opérateur choisi est la fonction t-conorme de Zadeh, qui retourne la valeur la plus grande en guise de degré d'appartenance final :

$$\mu_O^f(x, y) = \max(\mu_O^1(x, y), \mu_O^2(x, y), \dots, \mu_O^N(x, y))$$

le degré d'appartenance au fond restant :

$$\mu_F^f(x, y) = 1 - \mu_O^f(x, y)$$

Un pixel est considéré comme faisant partie du fond si son degré d'appartenance final μ_O^f est supérieur à 0,5. La figure 2 présente le résultat obtenu avec l'image de la figure 1(a). La comparaison avec l'image 1(b) montre bien que certains pixels ont été réintégrés au fond.

3 Résultats expérimentaux

Pour caractériser l'efficacité de la méthode proposée, nous avons retenu deux critères qualitatifs communément utilisés en segmentation d'image [LEV 85], le contraste et l'homogénéité, tels qu'ils sont définis dans [TAB 03]. Pour une image I divisée en R régions, le contraste est défini par :

$$C_I = \frac{\sum_{r \in I} |m_F - m_r|}{R} = |m_F - m_O|$$

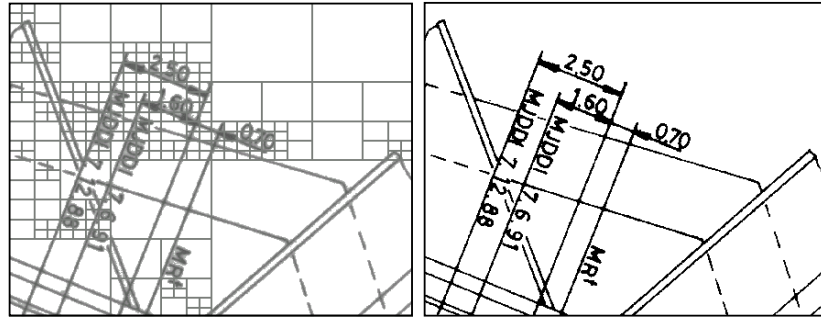
où m_F , m_O et m_r représentent respectivement les moyennes des valeurs des pixels appartenant au fond, à l'objet et à la région r . Plus la valeur de C_I est élevée, plus le contraste entre le fond et l'objet est grand.

Pour une image I dont l'objet O compte $\text{card}(O)$ pixels, l'homogénéité (de l'objet, donc) est définie par :

$$H_I = \sigma_O = \sqrt{\frac{\sum_{(x,y) \in O} (g(x, y) - m_O)^2}{\text{card}(O)}}$$

Moins la valeur de H_I est élevée, plus la région est homogène.

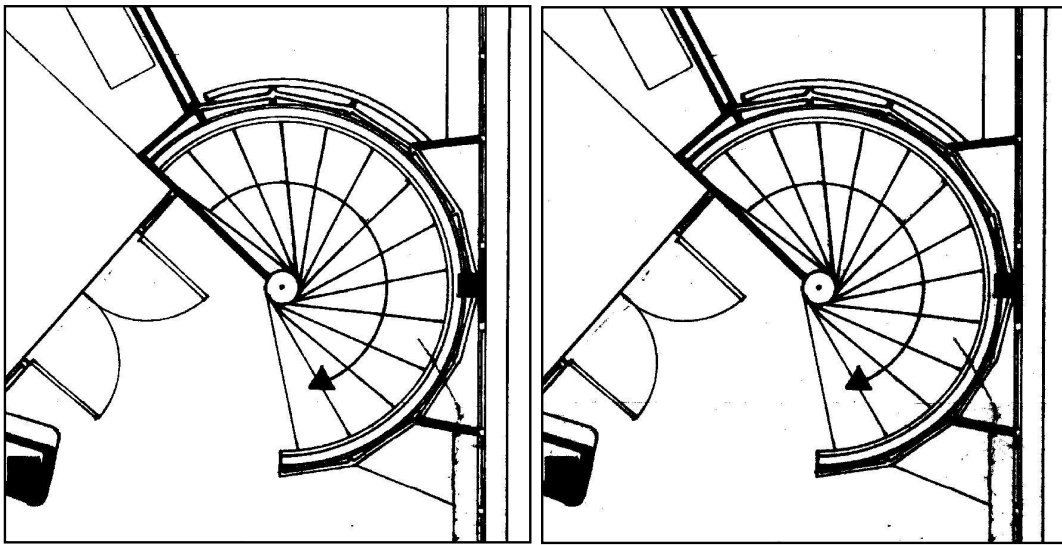
La figure 3 présente les résultats de la binarisation d'une même image, obtenus avec quatre méthodes différentes, celle



(a) découpage quaternaire

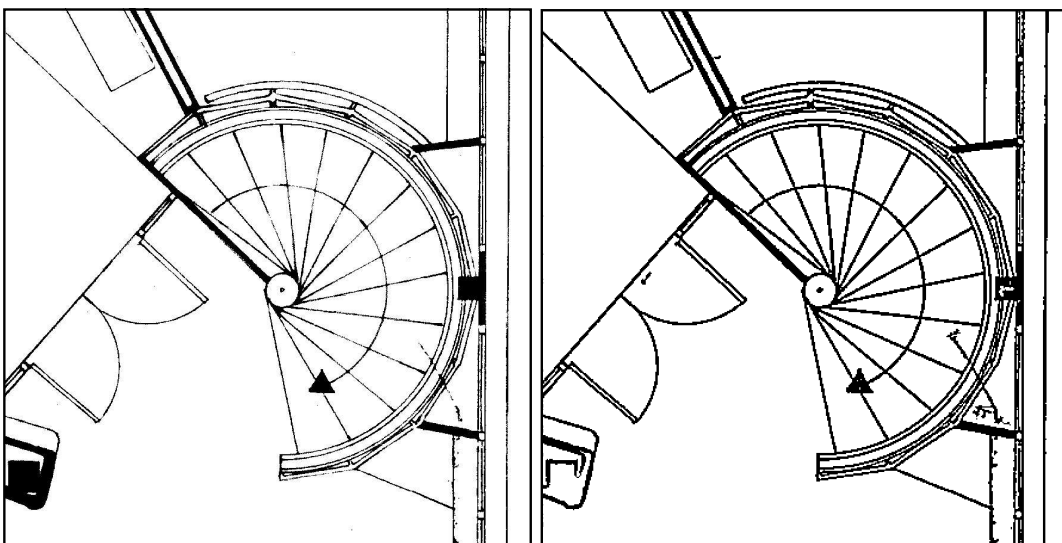
(b) image binaire finale

FIG. 2 – Deuxième étape : le seuillage adaptatif.



(a) Gadi et Benslimane [GAD 00]

(b) Otsu [OTS 79]



(c) Tabbone *et al.*

(d) Trier et Taxt [TRI 95]

FIG. 3 – Images binaires obtenues avec quatre méthodes différentes.

qui fait l'objet de ce papier (cf. Fig. 3(c)), celle de Gadi et Benslimane [GAD 00] (cf. Fig. 3(a)), dont s'inspire la précédente, celle de Otsu [OTS 79] (cf. Fig. 3(b)), qui est couramment utilisée pour la segmentation de caractères et qui a été retenue pour l'étape de seuillage global de notre méthode, et enfin celle de Trier et Taxt [TRI 95] (cf. Fig. 3(d)).

Le tableau suivant montre que notre méthode donne les meilleurs résultats en termes des critères de contraste et d'homogénéité que nous avons définis :

méthode	C_I	H_I
Cheng <i>et al.</i> [CHE 98]	123,23	43,94
Gadi et Benslimane [GAD 00]	124,82	44,26
Otsu [OTS 79]	130,19	39,19
Tabbone <i>et al.</i>	149,36	25,52
Trier et Taxt [TRI 95]	118,18	48,84

L'homogénéité des images finales fournies par la méthode de Gadi et Benslimane et la méthode de Trier et Taxt est moins bonne car, dans le premier cas, le fond n'est pas proprement segmenté tandis que, dans le second, des parties de l'objet sont affectées au fond.

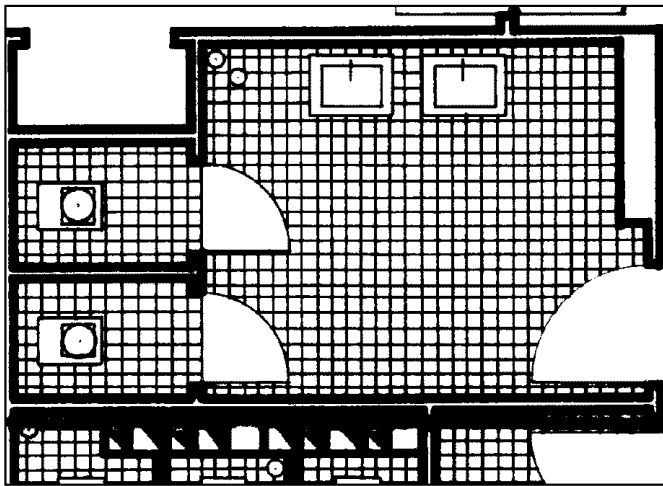
La figure 4 présente une autre série de résultats incluant la méthode de Cheng *et al.* [CHE 98], dont l'intérêt principal est de ne pas utiliser de seuil fourni *a priori*. On peut notamment constater que la méthode de Trier et Taxt n'est pas adaptée aux documents contenant des traits épais (cf. Fig. 4(b)). Elle utilise en effet un marquage des composantes connexes basé sur la polarité du Laplacien, qui n'est pas stable dans les zones bruitées, ce qui explique les trous à l'intérieur des triangles noirs en bas de l'image. Là encore, si l'on se réfère aux critères de contraste et d'homogénéité retenus, c'est notre méthode qui donne les meilleurs résultats (cf. Fig. 4(c)).

4 Conclusion

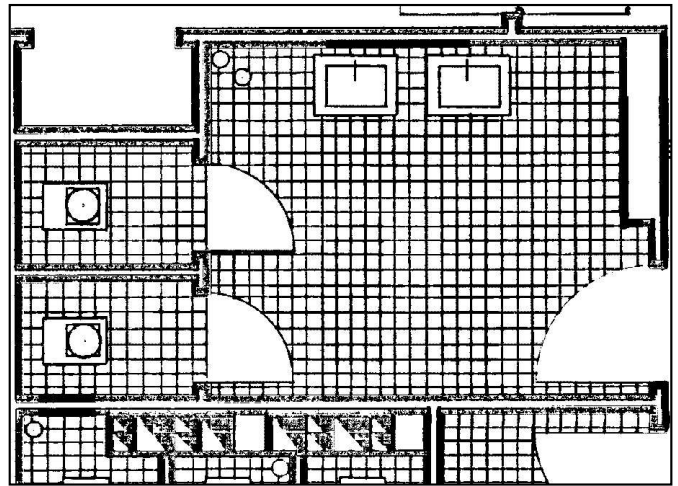
Les approches qui combinent des méthodes duales, locale et globale, semblent bien appropriées à la binarisation d'images de document graphique contenant à la fois des grandes et des petites composantes connexes. Le découpage de l'image en arbre quaternaire utilisé par la méthode décrite dans cet article facilite une telle combinaison. Toutefois, si les résultats obtenus sont prometteurs, la méthode nécessite encore quelques améliorations. D'une part, les hypothèses de normalité sur les distributions posées dans le paragraphe 2.2.1 ne sont pas vérifiées pour tous les documents. D'autre part, le principe du découpage d'une région à un niveau de l'arbre doit être assoupli par la prise en compte des propriétés locales de la région permettant de ne pas diviser systématiquement la région en quatre régions filles de tailles égales.

Références

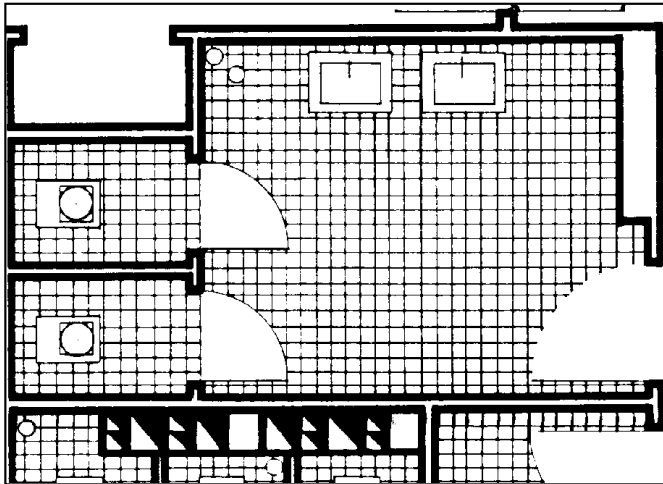
- [CHA 99] CHANG F., LIANG K. H., TAN T. M., HWANG W. L., Binarization of document images using Hadamard multiresolution analysis, *Proceedings of 5th International Conference on Document Analysis and Recognition (Bangalore, India)*, septembre 1999, pp. 157–160.
- [CHE 98] CHENG H. D., CHEN J. R., LI J., Threshold selection based on fuzzy c-partition entropy approach, *Pattern Recognition*, vol. 31, n° 7, 1998, pp. 857–870.
- [DAW 02] DAWOUD A., KAMEL M., Iterative model-based binarization algorithm for cheque images, *International Journal on Document Analysis and Recognition*, vol. 5, n° 1, 2002, pp. 28–38.
- [EIK 91] EIKVIL L., TAXT T., MOEN K., A fast adaptive method for binarization of document images, *Proceedings of 1st International Conference on Document Analysis and Recognition (Saint-Malo, France)*, vol. 1, 1991, pp. 435–443.
- [GAD 00] GADI T., BENSLIMANE R., Fuzzy hierarchical segmentation, *Traitement du Signal*, vol. 7, n° 1, 2000, pp. 59–67.
- [JAI 92] JAIN A. K., DUBUISSON M. P., Segmentation of X-ray and C-scan images of fiber reinforced composite materials, *Pattern Recognition*, vol. 25, n° 3, 1992, pp. 257–270.
- [KAM 93] KAMEL M., ZHAO A., Extraction of binary character/graphics images from grayscale document images, *CVGIP : Graphical Models and Image Processing*, vol. 55, n° 3, 1993, pp. 203–217.
- [KIM 02] KIM I. K., JUNG D. W., PARK R. H., Document image binarization based on topographic analysis using a water-flow model, *Pattern Recognition*, vol. 35, n° 1, 2002, pp. 265–277.
- [KUR 92] KURITA T., OTSU N., ABDELMALEK N., Maximum likelihood thresholding based on population mixture models, *Pattern Recognition*, vol. 25, n° 10, 1992, pp. 1231–1240.
- [LEV 85] LEVINE M. D., NAZIF A. M., Dynamic measurement of computer generated image segmentations, *IEEE Transactions on PAMI*, vol. 7, n° 2, 1985, pp. 155–164.
- [LIU 97] LIU Y., SRIHARI S. N., Document image binarization based on texture features, *IEEE Transactions on PAMI*, vol. 19, n° 5, 1997, pp. 540–544.
- [O'G 94] O'GORMAN L., Binarization and multithresholding of document images using connectivity, *CVGIP : Graphical Models and Image Processing*, vol. 56, n° 6, 1994, pp. 494–506.
- [OTS 79] OTSU N., A threshold selection method from grey-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-9, n° 1, 1979, pp. 62–66.
- [PAR 91] PARKER J. R., Graylevel thresholding in badly illuminated images, *IEEE Transactions on PAMI*, vol. 13, n° 8, 1991, pp. 403–411.
- [SAH 88] SAHOO P. K., SOLTANI S., WONG A. K. C., CHEN Y. C., A survey of thresholding techniques, *Computer Vision, Graphics and Image Processing*, vol. 41, n° 2, 1988, pp. 233–260.
- [SAU 00] SAUVOLA J., PIETIKÄINEN M., Adaptive document image binarization, *Pattern Recognition*, vol. 33, n° 2, 2000, pp. 225–236.



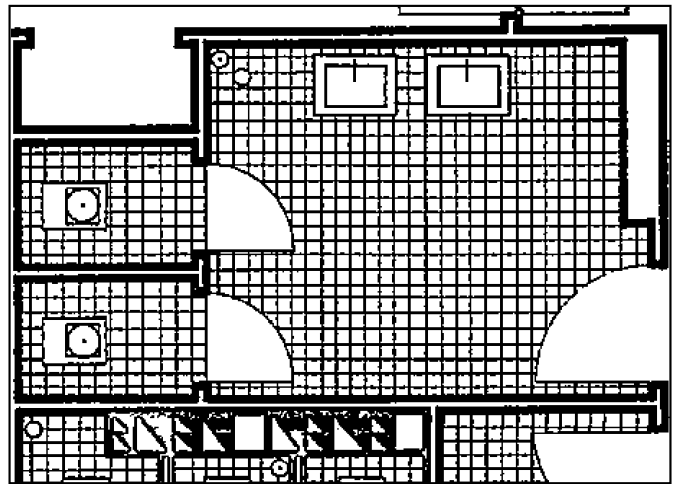
(a) Cheng *et al.* [CHE 98]



(b) Gadi et Benslimane [GAD 00]



(c) Tabbone *et al.*



(d) Trier et Text [TRI 95]

FIG. 4 – Images binaires obtenues avec quatre méthodes différentes.

[TAB 03] TABBONE S., WENDLING L., Multi-scale binarization of images, *Pattern Recognition Letters*, vol. 24, n° 1–3, 2003, pp. 403–411, Elsevier Science Publishers B.V. (North Holland).

[TAO 02] TAO W. B., TIAN J. W., LIU J., Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm, *Pattern Recognition Letters*, vol. 24, n° 16, 2002, pp. 3069–3078.

[TRI 95] TRIER Ø. D., TEXT T., Improvement of “integrated function algorithm” for binarization of document images, *Pattern Recognition Letters*, vol. 16, n° 3, 1995, pp. 277–283.

[YAN 00] YANG Y., YAN H., An adaptive logical method for binarization of degraded document images, *Pattern Recognition*, vol. 33, n° 5, 2000, pp. 787–807.